

I CLAIM:

- 1 1. A method of voice recognition, comprising the steps of:
2 organizing a plurality of speaker data points, representing a plurality of enrollment
3 speakers, into a data structure using high-dimensional vectors that represent
4 characteristics of enrollment voice samples from the enrollment speakers;
5 estimating a density of a subset of the plurality of speaker data points comprising the
6 approximate nearest neighbors to an unidentified voice sample from an
7 unidentified speaker; and
8 identifying the unidentified speaker based on one or more speaker data points most
9 closely matching the unidentified voice sample as indicated by the estimated
10 density.
- 1 2. The method of claim 1, wherein the step of estimating the density
2 comprises estimating a probability density function using Parzen windows to estimate the
3 probability density function.
- 1 3. The method of claim 1, wherein the step of estimating the density
2 comprises estimating the density based on a distance between individual speaker data points
3 within the subset of speaker data points
- 1 4. The method of claim 1, wherein the step of estimating the density
2 further comprises controlling the relative contributions of individual speaker data points
3 within the subset of speaker data points to the density based on a distance to a speaker data
4 point from the unidentified voice sample.
- 1 5. The method of claim 1, wherein the step of estimating the density
2 comprises estimating the density of the subset of speaker data points independent of
3 parametric distribution information related to the plurality of speaker data points.

1 6. The method of claim 1, wherein the data structure module organizes
2 the plurality of speaker data points such that a distance between individual speaker data
3 points is based on characteristic similarities between associated voice samples, the distance
4 measured in terms of one from the group containing: a Euclidean distance, a Minkowski
5 distance, and a Manhattan distance.

1 7. The method of claim 1, wherein the data structure comprises a kd-tree.

1 8. The method of claim 1, wherein the plurality of speaker data points
2 comprises a relatively large number of speaker data points.

1 9. The method of claim 1, further comprising a step of retrieving the
2 subset of speaker data points using an unidentified speaker data point from the unidentified
3 voice sample as an index into the plurality of speaker data points.

1 10. The method of claim 9, wherein the step of retrieving the subset of
2 speaker data points comprises retrieving approximate nearest neighbors to the unidentified
3 speaker data point, the approximate nearest neighbors comprising speaker data points within
4 a distance calculated as a function of a distance of an absolute nearest neighbor.

1 11. The method of claim 1, wherein the subset of speaker data points
2 includes more than one speaker data points associated with a common identification, and the
3 step of identifying the unidentified speaker accumulates a score for the common
4 identification.

1 12. The method of claim 1, further comprising extracting the high-
2 dimensional vectors from the enrollment voice samples and the unidentified voice sample.

1 13. The method of claim 1, wherein the step of identifying the unidentified
2 speaker comprises identifying the unidentified speaker as one of the enrollment speakers if
3 matching is within an error threshold.

1 14. The method of claim 1, wherein an enrollment voice sample and the
2 unidentified voice sample of a common speaker are text-independent.

1 15. A method of voice recognition, comprising the steps of:
2 retrieving a subset of speaker data points by using an unidentified speaker data point
3 as an index into a data structure comprising a plurality of speaker data points,
4 the subset of speaker data points representing approximate nearest neighbors
5 to the unidentified speaker data;
6 estimating a probability density function from a subset of the plurality of speaker data
7 points; and
8 identifying the unidentified speaker based on one or more speaker data points most
9 closely matching the unidentified voice sample as indicated by the probability
10 density function.

1 16. The method of claim 15, wherein the step of estimating the probability
2 density function comprises estimating the probability density function using Parzen windows
3 to estimate the probability density function.

1 17. A voice recognition system, comprising:
2 means for organizing a plurality of speaker data points, representing a plurality of
3 enrollment speakers, into a data structure using high-dimensional vectors that
4 represent characteristics of enrollment voice samples from enrollment
5 speakers;
6 means for estimating a density of a subset of the plurality of speaker data points
7 comprising the approximate nearest neighbors to an unidentified voice sample
8 from an unidentified speaker; and
9 means for identifying the unidentified speaker based on one or more speaker data
10 points most closely matching the unidentified voice sample as indicated by the
11 estimated density.

1 18. The system of claim 17, wherein the means for estimating uses Parzen
2 windows to estimate the density.

1 19. The system of claim 17, wherein the means for estimating estimates
2 the density based on a distance between individual speaker data points within the subset of
3 speaker data points.

1 20. The system of claim 17, wherein the means for estimating includes a
2 smoothing parameter to control the relative contributions of individual speaker data points
3 within the subset of speaker data points to the probability density function based on a
4 distance to a speaker data point from the unidentified voice sample.

1 21. The system of claim 17, wherein the means for estimating estimates
2 the density of the subset of speaker data points independent of parametric distribution
3 information related to the plurality of speaker data points.

1 22. The system of claim 17, wherein the means for organizing organizes
2 the plurality of speaker data points such that a distance between individual speaker data
3 points is based on characteristic similarities between associated voice samples, the distance
4 measured in terms of one from the group containing: a Euclidean distance, a Minkowski
5 distance, and a Manhattan distance.

1 23. The system of claim 17, wherein the means for organizing comprises a
2 kd-tree.

1 24. The system of claim 17, wherein the plurality of speaker data points
2 comprises a relatively large number of speaker data points.

1 25. The system of claim 17, further comprising means for retrieving the
2 subset of speaker data points uses an unidentified speaker data point from the unidentified
3 voice sample as an index into the plurality of speaker data points.

1 26. The system of claim 25, wherein the means for retrieving the subset of
2 speaker data points retrieves approximate nearest neighbors to the unidentified speaker data
3 point, the approximate nearest neighbors comprising speaker data points within a distance
4 calculated as a function of a distance of an absolute nearest neighbor.

1 27. The system of claim 17, wherein the subset of speaker data points
2 includes more than one speaker data points associated with a common identification, and the
3 identification module accumulates a score for the common identification.

1 28. The system of claim 17, further comprising a means for extracting the
2 high-dimensional vectors from voice samples.

1 29. The system of claim 17, wherein the means for identifying identifies
2 the unidentified speaker as one of the enrollment speakers if matching is within an error
3 threshold.

1 30. The system of claim 17, wherein an enrollment voice sample and the
2 unidentified voice sample of a common speaker are text-independent.

1 31. A computer program product, comprising:
2 a computer-readable medium having computer program instructions and data
3 embodied thereon for voice recognition, comprising the steps of:
4 organizing a plurality of speaker data points, representing a plurality of
5 enrollment speakers, into a data structure using high-dimensional
6 vectors that represent characteristics of enrollment voice samples from
7 the enrollment speakers;
8 estimating a density of a subset of the plurality of speaker data points
9 comprising the approximate nearest neighbors to an unidentified voice
10 sample from an unidentified speaker; and
11 identifying the unidentified speaker based on one or more speaker data points
12 most closely matching the unidentified voice sample as indicated by
13 the estimated density.

1 32. The computer program product of claim 31, wherein the step of
2 estimating the density comprises estimating a probability density function using Parzen
3 windows to estimate the probability density function.

1 33. The computer program product of claim 31, wherein the step of
2 estimating the density comprises estimating the density based on a distance between
3 individual speaker data points within the subset of speaker data points

1 34. The computer program product of claim 31, wherein the step of
2 estimating the density further comprises controlling the relative contributions of individual
3 speaker data points within the subset of speaker data points to the probability density
4 function based on a distance to a speaker data point from the unidentified voice sample.

1 35. The computer program product of claim 31, wherein the step of
2 estimating the density comprises estimating the probability density function of the subset of
3 speaker data points independent of parametric distribution information related to the plurality
4 of speaker data points.

1 36. The computer program product of claim 31, wherein the data structure
2 module organizes the plurality of speaker data points such that a distance between individual
3 speaker data points is based on characteristic similarities between associated voice samples,
4 the distance measured in terms of one from the group containing: a Euclidean distance, a
5 Minkowski distance, and a Manhattan distance.

1 37. The computer program product of claim 31, wherein the data structure
2 comprises a kd-tree.

1 38. The computer program product of claim 31, wherein the plurality of
2 speaker data points comprises a relatively large number of speaker data points.

1 39. The computer program product of claim 31, further comprising a step
2 of retrieving the subset of speaker data points using an unidentified speaker data point from
3 the unidentified voice sample as an index into the plurality of speaker data points.

1 40. The computer program product of claim 39, wherein the step of
2 retrieving the subset of speaker data points comprises retrieving approximate nearest
3 neighbors to the unidentified speaker data point, the approximate nearest neighbors
4 comprising speaker data points within a distance calculated as a function of a distance of an
5 absolute nearest neighbor.

1 41. The computer program product of claim 31, wherein the subset of
2 speaker data points includes more than one speaker data points associated with a common
3 identification, and the identification module accumulates a score for the common
4 identification.

1 42. The computer program product of claim 31, further comprising
2 extracting the high-dimensional vectors from the enrollment voice samples and the
3 unidentified voice sample.

1 43. The computer program product of claim 31, wherein the step of
2 identifying the unidentified speaker comprises identifying the unidentified speaker as one of
3 the enrollment speakers if matching is within an error threshold.

1 44. The computer program product of claim 31, wherein an enrollment
2 voice sample and the unidentified voice sample of a common speaker are text-independent.